Lecture Notes in Computer Science 801

Jan-Olof Eklundh (Ed.)

Computer Vision – ECCV '94

Third European Conference on Computer Vision
Stockholm, Sweden, May 1994
Proceedings, Volume II

Springer-Verlag

Jan-Olof Eklundh (Ed.)

Computer Vision – ECCV '94

Third European Conference on Computer Vision
Stockholm, Sweden, May 2-6, 1994
Proceedings, Volume II

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Contents

Planning the Optimal Set of Views Using the Max-Min Principle*

Jasna Maver, Aleš Leonardis, and Franc Solina

Computer Vision Laboratory, Faculty of Electrical Engineering and Computer Science
University of Ljubljana, Tržaška 25, 61001 Ljubljana, Slovenia
E-mail: Jasna.Maver@fer.uni-lj.si

Abstract. We present an approach to a typical task of active sensing, namely how to successively position the sensor in order to accomplish a given task. The approach is data driven since it is only the data that provides the information for planning the next views, and no a priori knowledge about the constituents of the scene is necessary. We applied our approach to the task of determining the 3-D coordinates of vertices of object silhouettes in the image under orthographic projection.

1 Introduction

It has long been realized that information obtained from a single viewpoint might not be sufficient for successfully accomplishing a task and that additional views are necessary [1,2,4,5,6].

In this paper we present an active sensor system which tends to accomplish a given task by acquiring a minimal number of images. The principle which guides the next-view planning is the *Max-Min principle*. The viewing direction from which the first image is acquired is arbitrary or predefined. To select the new viewing direction for image acquisition, the sensor system must estimate how much of the yet unknown data necessary to accomplish the task can be acquired from each possible viewing direction. This estimate is based only on the information acquired in the previous images. When there is insufficient information the worst-case situation is assumed. Thus for each viewing direction the minimal amount of data that can be obtained is anticipated, i.e., the utilization of the minimum principle. To select the best viewing direction for the next view we propose two different strategies:

1. From all possible viewing directions select the one which gives under the assumption of worst-case situation for the unknown data the maximal amount of new information.
2. First compute the necessary viewing directions from which all the necessary data to accomplish the task can be acquired, and then select among these viewing directions the one which maximizes the amount of new information.

* This research has been supported in part by The Ministry for Science and Technology of The Republic of Slovenia (Project P2-1122).

Lecture Notes in Computer Science, Vol. 801
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

While the first strategy can be compared to a greedy algorithm which tries to obtain at each step the maximal information regardless of the final number of views, the second strategy with its look-a-head capability takes into account a more global view.

2 Geometry of the Viewer

The scene is defined in a rectangular coordinate system (X, Y, Z). The viewer's

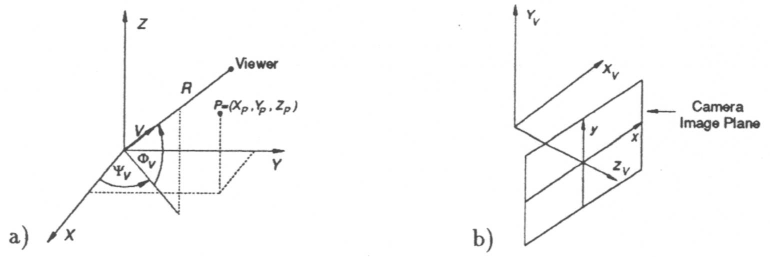

Fig. 1. (a) Viewing direction **V**, (b) Camera image plane.

position is on a half sphere with radius R (Fig. 1(a)). The unit vector **V** pointing toward the viewer is given by $\mathbf{V}^T = \left[\cos\Psi_V \cos\Phi_V, \sin\Psi_V \cos\Phi_V, \sin\Phi_V\right]$. $\Psi_V \in [0, 2\pi)$ denotes the azimuth and $\Phi_V \in [0, \frac{\pi}{2}]$ the elevation. The viewer's coordinate system (X_V, Y_V, Z_V) is located in the center of the scene coordinate system. The coordinates of a point $P = (X_P, Y_P, Z_P)$ are in the viewer's coordinate system given as $(\mathbf{X}_V = \mathbf{Z} \times \mathbf{V}, \mathbf{X}_V \times \mathbf{Y}_V = \mathbf{V}, \mathbf{Z}_V = \mathbf{V})$

$$\begin{bmatrix} X_{V_P} \\ Y_{V_P} \\ Z_{V_P} \end{bmatrix} = \begin{bmatrix} -\sin\Psi_V & \cos\Psi_V & 0 \\ -\cos\Psi_V \sin\Phi_V & -\sin\Psi_V \sin\Phi_V & \cos\Phi_V \\ \cos\Psi_V \cos\Phi_V & \sin\Psi_V \cos\Phi_V & \sin\Phi_V \end{bmatrix} \begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} . \tag{1}$$

The camera image plane is parallel to the viewer's X_V-Y_V plane where the image axes (x, y) are defined so as to be parallel to the axes (X_V, Y_V) (Fig. 1(b)). Z_V is orthogonal to the image plane. We assume that the imaging is performed under the orthographic projection:

$$x = X_V, \qquad y = Y_V . \tag{2}$$

3 Task Definition

A set of objects lies on the horizontal base surface of the scene. Without loss of generality we can assume that the first image is acquired from the viewing direction $\mathbf{V}_1 = (0, 0, 1)$, i.e., $\Psi_{V_1} = -\frac{\pi}{2}, \Phi_{V_1} = \frac{\pi}{2}$. In the image obtained from the first viewing direction the vision system extracts the silhouettes of the objects[2].

[2] We assume that our vision system can distinguish whether an image point is a projection of the object or a projection of the support plane.

The task that we study can now be formulated as: *In the image of the first view determine the 3-D coordinates of the vertices on the detected silhouettes.*

4 Geometrical Information in the First two Views

To design an efficient algorithm we first have to answer the following two questions: 1. Which geometrical information of object points acquired in the image of the first view can be extracted from the image? 2. Which geometrical information can be obtained about these points from the second view?

First view: Let $p = (x_p, y_p)$ be a projection of an object point $P = (X_P, Y_P, Z_P)$. Eqs. (1) and (2) yield: $X_P = X_{V_P} = x_p$, $Y_P = Y_{V_P} = y_p$. The coordinate that remains unknown after the first view is Z_P coordinate.

Second view: Let the second viewing direction be (Ψ_{V_2}, Φ_{V_2}). Object points on the line $y_{\Psi_{V_2},n} = \tan(\Psi_{V_2})x + n$ in the image of the first view have their corresponding points on the line $x = n\cos(\Psi_{V_2})$ in the image of the second view (if they are not occluded) [3]. The assumption that the point p on the line $y_{\Psi_{V_2},n}$ (Fig. 2) projects into the point $(n\cos(\Psi_{V_2}), y_{max_2})$ gives the upper bound of the height of the scene at the location (X_P, Y_P). We denote the computed bound as Z_{max}. After the second view we know for each object point p its (X_P, Y_P) coordinates and the upper bound on its Z coordinate $Z_P \leq Z_{max}(X_P, Y_P)$.

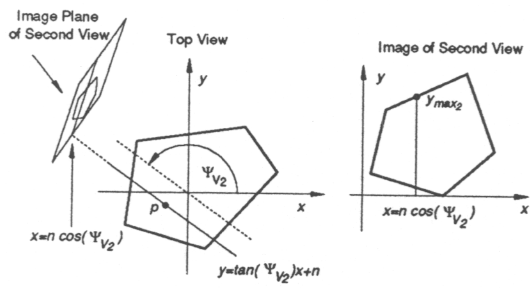

Fig. 2. Computation of Z_{max}.

5 Computation of Azimuth and Elevation Angle

From the first view we get (X, Y) coordinates of the vertices. All the following views are selected to determine their Z coordinates. Two conditions must be met to determine the Z coordinate of a vertex from the next view: 1. The vertex must not be occluded by any part of the scene. 2. The vertex must be uniquely determined in the image.

For each convex vertex v_i in the image of the first view we compute the set of possible viewing directions—the *azimuth angle* $\Psi(v_i)$ and for the selected azimuth the *elevation angle* $\Phi(v_i)$ [3] from which the viewer can uniquely determine 3-D coordinates of the vertex V_i.

Possible viewing directions for the second view: After the first view the height of the scene remains completely undetermined, therefore to compute $\Psi(v_i)$ and $\Phi(v_i)$ the following two assumptions are made for each azimuth value Ψ_{V_2}:

1. Vertex V_i is occluded if there is an object point on the line $y_{\Psi_{V_2},n}$ in the image of the first view in front of v_i relative to the viewer[3].
2. The corresponding point of v_i can not be uniquely determined if there is an object point on the line $y_{\Psi_{V_2},n}$ in the image of the first view behind v_i relative to the viewer.

Using these two assumptions we determine for each vertex v_i the minimal set of viewing directions. The slopes of the lines $y_{\Psi_{V_2},n}; n \in \Re, \Psi_{V_2} \in [0, 2\pi)$ which pass only through the vertex v_i define the set of all azimuth values—azimuth angle $\Psi(v_i)$. The elevation angle, which is equal for all vertices, is: $\Phi(v_i) = [0, \frac{\pi}{2})$ for any $\Psi_{V_2} \in \Psi(v_i)$. The point in the image of the second view corresponding to v_i is the point on the line $x = n \cos \Psi_{V_2}$ with the largest y coordinate [3].

Possible viewing directions for the third and all subsequent views: Knowing the Z_{max} coordinate (which is updated after each view) for each pair (X, Y) of the scene, we compute new azimuth and elevation angles for vertices v_i.

Let the two points p_1 and p_2 lie on the line $y_{\Psi_{p_1,p_2},n}$ in the image of the first view (Fig. 3). p_1 and p_2 are the projections of the surface points P_1 and P_2, respectively. Let the point P_1 be in front of the point P_2 relative to the viewer. The scene at the location (X_{P_1}, Y_{P_1}) can be occupied for $0 \le Z \le Z_{max}(X_{P_1}, Y_{P_1})$ and at the location (X_{P_2}, Y_{P_2}) for $0 \le Z \le Z_{max}(X_{P_2}, Y_{P_2})$.

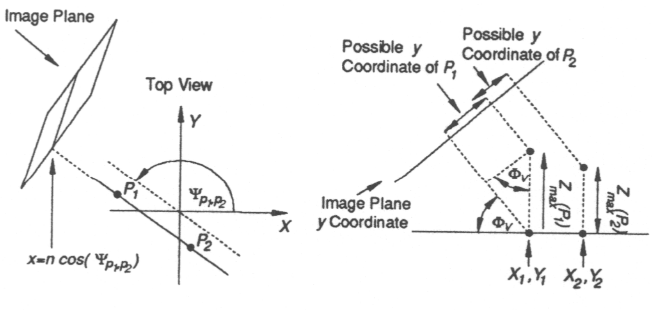

Fig. 3. Computation of the minimal viewing elevation.

The y coordinates of P_1 and P_2 lie on the intervals (Fig. 3):
$y((X_{P_1}, Y_{P_1}, 0), \Psi_{p_1,p_2}, \Phi_V) \le y_{p_1} \le y((X_{P_1}, Y_{P_1}, Z_{max}(X_{P_1}, Y_{P_1})), \Psi_{p_1,p_2}, \Phi_V)$,
$y((X_{P_2}, Y_{P_2}, 0), \Psi_{p_1,p_2}, \Phi_V) \le y_{p_2} \le y((X_{P_2}, Y_{P_2}, Z_{max}(X_{P_2}, Y_{P_2})), \Psi_{p_1,p_2}, \Phi_V)$,
respectively. The two intervals do not overlap if

$$\Phi_V > \arctan\left(\frac{Z_{max}(X_{P_1}, Y_{P_1})}{\sqrt{(Y_{P_1} - Y_{P_2})^2 + (X_{P_1} - X_{P_2})^2}}\right) = \Phi_{min}(p_1, p_2) \ . \tag{3}$$

[3] v_i lies on the line $y_{\Psi_{V_2},n}$.

If p_2 is a vertex then for the viewing direction $(\Psi_{p_1,p_2}, \Phi_V > \Phi_{min}(p_1, p_2))$ the scene at the location (X_{P_1}, Y_{P_1}) does not occlude P_2. If p_1 is a vertex then for $(\Psi_{p_1,p_2}, \Phi_V > \Phi_{min}(p_1, p_2))$, we can distinguish the projection of P_1 from the projections of the points at (X_{P_2}, Y_{P_2}) [3]. Taking into account Z_{max} for each scene location the new azimuth and elevation angles can be computed.

6 Determining the Next Viewing Direction

For the next viewing direction we must determine the azimuth and the elevation value. The selection of azimuth has a dominant role.

First we build a histogram $\mathcal{H}(\psi)$ which shows the number of vertices that can be determined for any azimuth value $\psi \in [0, 2\pi)$.

One possible solution is to select as the next azimuth value the one from which the largest number of vertices can be determined, i.e., ψ for which $\mathcal{H}(\psi)$ has a global maximum. Another possible solution to the problem is first to analyze the histogram $\mathcal{H}(\psi)$ and find the necessary number of viewing directions from which we can compute the Z coordinate of all the vertices, and then select among them the one from which the largest number of vertices can be determined. Due to the lack of space the reader is referred to [3].

7 Experimental Results

A scene consists of two rectangloids and two pyramids. In the image of the first view (Fig. 4(a)) we locate 16 vertices. For each vertex we compute the azimuth angle and build the histogram (Fig. 4(b)). For Ψ_{V_2} we choose the value from one of the histogram maxima: $\Psi_{V_2} = 270,5°$. Φ_{V_2} can be any value from the interval $[0, \frac{\pi}{2})$. We choose $\Phi_{V_2} = 76.5°$. From this view we can determine the Z coordinates of six vertices marked by circles, as shown in Fig. 4(c). Fig. 5(a) shows the image taken from the second view. Fig. 5(b) depicts the y_{max_2} coordinates from which we can get the Z_{max} coordinates. We compute new azimuth angles for the remaining 10 vertices. Fig. 5(c) depicts the new histogram. Two additional

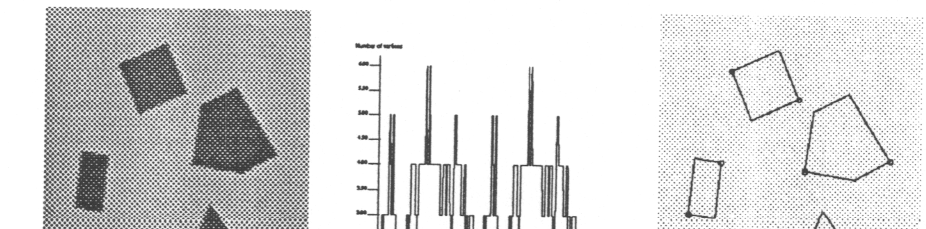

Fig. 4. (a) Image of the scene taken from the first view, (b) Histogram showing the number of vertices for which the Z coordinate can be determined with respect to the azimuth value, (c) Six selected vertices.

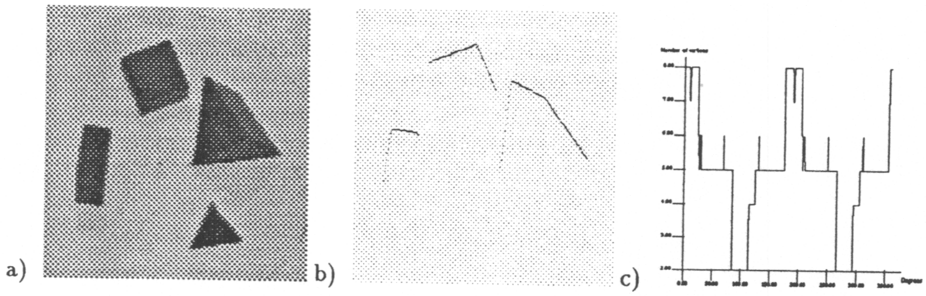

Fig. 5. (a) The image of the scene taken from the second viewing direction, (b) y_{max_2} coordinates, (c) Histogram of remaining 10 vertices.

images must be acquired to accomplish the task: one azimuth value must be selected from the interval $(175°, 189°)$ or $(355°, 9°)$, another from the interval $(27°, 85°)$ or $(207°, 265°)$. We choose as the third azimuth value $\Psi_{V_3} = 2°$. For the vertices v_i which are selected by this azimuth we compute the minimal elevation $\Phi_{min}(v_i)$. Elevation must be $\Phi_{V_3} > \max_i(\Phi_{min}(v_i))$. We compute: $\Phi_{V_3} > 67°$.

8 Conclusion

In this paper we explored the issue of selecting the optimal set of views from which the viewer can determine the 3-D coordinates of convex vertices of objects silhouettes. However, the information acquired in the set of images is much richer—we also get the limits of the height of each object point in the scene, Z_{max}[4]. This information can be helpful in many robotic tasks to avoid collisions of robot arm with the objects in the scene. The method can also be extended to non-polyhedral scenes. In that case the foci of attention can be the convex parts of object silhouettes.

References

[1] J. Aloimonos and A. Bandyopadhyay, "Active Vision," In *Proc. of the IEEE 1 st Int. Conf. on Computer Vision*, pp. 35–54, June 1987.

[2] R. Bajcsy, "Active Perception", In *Proc. of the IEEE*, vol.76, no.8, pp. 996–1005, August 1988.

[3] J. Maver, *Planning the Next View Using the Max-Min Principle*, University of Ljubljana, Faculty of EE and CS, Tech. report LRV-93-4, 1993.

[4] J. Maver and R. Bajcsy, "Occlusions as a Guide for Planning the Next View," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol.15 no.5, pp. 417–433, 1993.

[5] D. Wilkes and J. K. Tsotsos, "Active Object Recognition," In *Proc. of the CVPR'92*, pp. 136-141, 1992

[6] J. Y. Zheng, F. Kishino, Q. Chen, and S. Tsuji, "Active Camera Controlling for Manipulation," In *Proc. of the CVPR'91*, pp. 413–418, 1991.

[4] That is true even though the point is occluded in all subsequent views.